STATISTICAL LABORATORY #1

When Can Confounding Reverse an Association?

Investigating the role of a third variable

Hosted by Nicola Orsini

Department of Global Public Health

November 6, 2026

Why bother?

A stakeholder has mortality data from several pilot studies and is considering whether to stop use of a drug.

What should the stakeholder conclude from the data?

Your task is to determine what happens in the whole sample, what happens within age groups, and why the answers may differ.

The recommendation matters because a decision based on the wrong comparison could withdraw a useful drug or retain a harmful one.

Study and variables

Study design

A pilot observational cohort. Age and drug use were recorded at baseline, and mortality was assessed within five years.

Available data

One overall count table and two tables stratified by age. In every table, Y is in the rows and X is in the columns.

Variables

Z — Age

$0 =$ younger, $1 =$ older

X — Drug use

$0 =$ no drug, $1 =$ drug

Y — Five-year mortality

$0 =$ alive, $1 =$ died

Questions for the pilot study

1. What is the mortality risk difference between the drug groups in the whole sample?
2. What are the drug risk differences within the younger and older groups?
3. Does the direction of the overall result differ from both age-specific results?
4. Which drug group contains a larger proportion of older participants?
5. How different is mortality between older and younger participants within each drug group?
6. What should the stakeholder conclude, and what remains uncertain?

The two-hour laboratory

Time	Activity
10:00–10:10	The challenge
10:10–10:40	Pilot study
10:40–10:50	Gallery wall
10:50–11:00	Mechanism and funding
11:00–11:40	New evidence
11:40–12:00	Reveal and discussion

Unexpected results are part of the investigation.

What goes on the wall

Report only selected quantities and conclusions. Do not copy the count tables.

Heading	Include
Group	Group number and sample size
Overall	RD_{overall}
Conditional drug effect	$RD_{XY\mid Z=0}$ and $RD_{XY\mid Z=1}$
Age composition	$q_1 = P(Z = 1 \mid X = 1)$ and $q_0 = P(Z = 1 \mid X = 0)$
Conditional age effect	$RD_{ZY\mid X=0}$ and $RD_{ZY\mid X=1}$
Pattern	Reversal: yes or no, with one explanatory sentence
Conclusion	Explanation and recommendation to the stakeholder

Round probabilities and risk differences to three decimal places.

Before you start

- Calculate the risks from the counts and show enough working for another group to follow.
- Describe the overall and age-specific results before trying to explain them.
- Separate what this sample shows from what you would expect in repeated samples.
- In the recommendation, distinguish the observed associations from the evidence needed for a decision about the drug.

Should the stakeholder stop use of the drug?

Let the tables guide your answer.